SCHWA REDUCTION IN LOW-PROFICIENCY L2 SPEAKERS: LEARNING AND GENERALIZATION

Lisa Morano[a], Mirjam Ernestus[a,b], and Louis ten Bosch[a]
[a]Radboud University; [b]Max Planck Institute for Psycholinguistics; Nijmegen, the Netherlands
l.morano@let.ru.nl; m.ernestus@let.ru.nl; l.tenbosch@let.ru.nl

ABSTRACT

This paper investigated the learnability and generalizability of French schwa alternation by Dutch low-proficiency second language learners. We trained 40 participants on 24 new schwa words by exposing them equally often to the reduced and full forms of these words. We then assessed participants' accuracy and reaction times to these newly learnt words as well as 24 previously encountered schwa words with an auditory lexical decision task. Our results show learning of the new words in both forms. This suggests that lack of exposure is probably the main cause of learners' difficulties with reduced forms. Nevertheless, the full forms were slightly better recognized than the reduced ones, possibly due to phonetic and phonological properties of the reduced forms. We also observed no generalization to previously encountered words, suggesting that our participants stored both of the learnt word forms and did not create a rule that applies to all schwa words.

Keywords: L2 acquisition, generalization, speech comprehension, French schwa reduction.

1. INTRODUCTION

Reduced word pronunciation forms (*i.e.* forms with weakened or absent segments compared to the words' citation forms) are highly pervasive in everyday conversations [1]. Importantly, while words' reduced forms are easily understood by natives, who are barely aware of them, second language (L2) learners struggle comprehending them [2]. One of the possible reasons for low proficiency learners' problems with comprehending reduced pronunciation forms may be related to the classroom environment itself where L2 learners are hardly exposed to reduced forms.

In this paper, we address two research questions. First, we investigate whether L2 beginning learners recognize the reduced and full forms of words equally well if they have heard both forms of these words equally often. If they do, they may have learned that, for these words, two pronunciation forms are possible, or they may have formed a generalization which they also apply to other words. This brings us to our second research question: can learners generalize knowledge about reduction processes observed in newly learnt words to previously encountered words. For instance, if learners of English are familiarized with the variations /ˈæmbjələns/-/ˈæmbjləns/ for *ambulance* and /ˈæbsəlut/-/ˈæbslut/ for *absolute*, can they then also easily understand /ˈkæθlɪk/ for *catholic*?

We are not the first ones to investigate whether learners of a (artificial) language can generalize a sound pattern from one word to another word (e.g. [3]). To our knowledge, all generalization studies investigated generalization of a newly acquired pattern to unknown new words. While learning a foreign language, however, learners may have to generalize a pattern to words they already know. For instance, they may already know the full form of the word *catholic* before they learn that *absolute* can be pronounced as both /ˈæbsəlut/ and /ˈæbslut/. We investigated whether, given successful implicit learning of reduced forms of some words, learners can generalize a speech reduction pattern to previously encountered words.

To address these questions, we trained and tested Dutch beginning learners on French alternating schwa words (e.g. *pelouse* 'lawn' can be pronounced with schwa /pəluz/ or without /pluz/). Schwa reduction is very frequent in French, even in more formal speech [4]. It also frequently occurs in Dutch, but mostly results in a shortened rather than completely absent schwa, and rarely occurs in formal speech [5].

We first trained participants on the full and reduced forms of new schwa words. Subsequently we tested them in an auditory lexical decision task on the full and reduced forms of these new words as well as previously encountered schwa words.

2. EXPERIMENT

2.1. Participants

We tested 40 Dutch native speakers between 18 and 24 years old (mean 20). All studied French in high school for maximally three years and were paid for their participation. None reported any hearing impairment. Five were left-handed and eight were male. Prior to the experiment, the participants rated their skills in French on a zero (very bad) to five (very good) point scale. They scored on average 1.5 for reading, 1.3 for listening, 0.9 for writing, and 0.9 for speaking.

2.2. Materials

To better hide the aim of the experiment to the participants, the participants learned two types of words: 24 words with and 24 words without schwa in their initial syllables. The trained words with schwa (henceforth Trained Target words, TTs) were bisyllabic real French words (e.g. *semoule* 'semolina'), starting with only one consonant, and considered by native speakers as equally acceptable in their reduced (e.g. /smul/) and full (e.g. /səmul/) forms [6]. Absence of schwa in these words does not lead to assimilation in the resulting consonant sequence. The trained words without schwa (Trained Fillers, TFs) consisted of twelve monosyllabic and twelve bisyllabic words. All 48 words were chosen from two beginners' textbooks frequently used in the Netherlands [7] and all were depictable. We selected 48 pictures to represent the TTs and the TFs from three free-of-rights picture databases [8].

We also selected 24 schwa words likely to be known by these participants (Non-Trained Target words, NTTs): 13 schwa words from the same beginners' textbooks previously mentioned and eleven cognate schwa words (either in Dutch or in English, e.g. *menu* /məny/ 'menu').

In the lexical decision experiment, the 24 TTs, 24 TFs and 24 NTTs were intermixed with 24 real words (non-Trained Fillers; NTFs), which were selected in the same way as the TFs, as well as 96 phonotactically legal pseudowords. We created the pseudowords by reusing the first syllable or consonant cluster of every bi- and monosyllabic real words. These pseudowords matched the real words in number of syllables and Consonant-Vowel (CV) structure. For instance, *la remarque* (/larəmark/) formed the basis for *le recombre* (/lərəkõbr/). Finally, we selected six more non-schwa real words and created six pseudowords for practice trials.

We created five different pseudo-randomizations of the trials, respecting the following constraints: 1) the very first stimulus and the first stimulus after the break in the middle of the task were not TTs; 2) a TT was never followed by another TT, and 3) no more than eight real words or pseudowords occurred sequentially. We included two tokens of some words (the 24 TTs, 6NTFs, 6TFs and 36 pseudowords), which led to 264 trials per list. In each list, approximately half of all stimuli (including the non-schwa words) were reduced. Each list was assigned to eight participants.

2.3. Recordings and speakers

A female Dutch native speaker from The Hague recorded the Dutch translations of the 48 trained words. A male French native speaker from Paris (speaker Fr1) recorded all 204 French word types (real and pseudowords) preceded by the definite determiner (*le* or *la*). He first produced them in a careful speech style (by enunciating clearly) and then in a casual speech style (by reducing pronunciation effort, and, for the schwa words and pseudowords, by dropping the schwa). A female French native speaker from Lyon (speaker Fr2) recorded the 48 training words. She listened to the recordings from speaker Fr1 and repeated after him. We selected the two best tokens for each speech style from each speaker. The stimuli were recorded in a sound attenuated booth at a 44.1kHz sampling rate and 16-bit resolution on a mono channel. The recordings were manually spliced in separate sound files and normalized for amplitude with Praat [9].

All word tokens recorded in casual speech style were shorter than the corresponding tokens recorded in careful speech style, for both speakers and for all word types (on average 121 ms shorter for schwa words and 126 ms for non-schwa words). Table 1 shows the average durations of the reduced and full tokens of the schwa words.

Table 1: Average stimuli duration in milliseconds per French native speaker, per stimuli type per speech style.

Speaker	TTs		NTTs	
	reduced	full	reduced	full
Fr1 - Test	579	723	568	664
Fr2 - Training	597	732	n.a.	n.a.

2.4. Procedure

The participants were trained and tested individually in sound attenuated booths. The stimuli were always presented over headphones at a comfortable listening volume. In the training phase, the pictures and the French words were presented with PsychoPy [10] and in the test phase the words were presented with Presentation [11].

The participants were first familiarized with the 48 pictures: they saw each picture once while hearing the corresponding Dutch word. Then they saw the 48 pictures one by one again and heard for each picture four tokens of the corresponding French word in a sequence (speaker Fr2): two reduced and two full tokens in random order. The participants were asked to learn the words. Afterwards, the participants had six rounds of training in which, upon hearing a French word, they had to click on the corresponding picture on the screen. In every round, the 48 words were played once, in random order and randomly uttered either in reduced or full form. The number of pictures displayed on the screen increased at each practice round, from four to six, and was then maintained to six pictures. If participants clicked on the correct picture, a green check mark briefly appeared on the picture and the next trial was initiated; otherwise, a red cross appeared and the participant could click on another picture. At the end of each practice round, the participant's accuracy was displayed on the screen. By the end of the training, the participants had heard each TT and each FT five times reduced and five times in full.

The participants were then immediately tested with an auditory lexical decision task. They were instructed to press the 'yes' button of a button box with their dominant hand if the word they heard was a real word in French and to press the 'no' button with their other hand if they thought it was not a real French word or if they had any doubt. The next stimulus started one second after the participant's response or 3.5s after the onset of the previous stimulus if the participant did not press any button. The task started with 16 practice trials (4 practice words were repeated) after which the experiment leader checked if the participant understood the instructions and that the volume was set to a comfortable level. There was a brief pause after the first half of the stimuli. In order to discourage guessing and to reward participants' learning efforts, accuracy feedback (in percentages of correct answers) was displayed after the practice trials, just before the break, and at the end of the task.

3. RESULTS

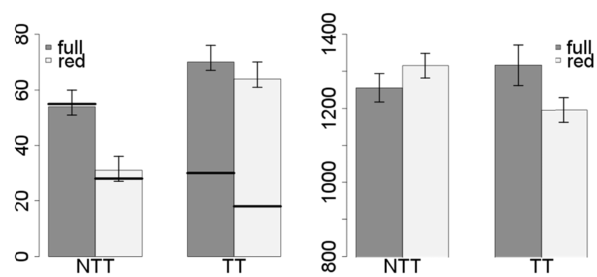
The participants' high accuracy in recognizing the non-schwa words (97% for the TFs, 84% for the NTFs) and in rejecting the pseudowords (81%) shows that they took the task seriously. Their lowest accuracy scores were for the schwa words: 67% for the TTs and 43% for the NTTs.

We analysed the participants' accuracies and reaction times (RTs, measured from word onset) to the first tokens of the TTs and NTTs in the lexical decision task. Accuracies were analysed by means of generalized mixed effects models [12], and RTs to correct trials within 2.5 standard deviations from the targets' means by means of mixed effects regression models [13]; both with the software R [14]. Prior to analysis, all RTs and word durations were log-transformed. Only predictors which showed significant effects were retained in the final models. Interactions were tested for the predictor of interest. We used item and participant as random effects. Random slopes were tested for all fixed effects. The final RT models were refitted after removal of the outliers: RTs with residual standard errors more than 2.5 standard deviation units were excluded from the datasets of the final statistical models.

3.1. Accuracies to Targets words (TTs and NTTs)

We compared participants' accuracies to the target words as a function of reduction and training (*cf.* Figure 1, left, and Table 2). A group of twelve control participants with the same proficiency level as our participants performed the lexical decision task without any training. The accuracy scores of these participants are also plotted on the Fig.1 left.

Figure 1: Accuracy scores in percentages of correct answers (left), and RTs to correct answers in milliseconds (right) for the trained (TT) and non-trained (NTT) target words' first occurrence, split by pronunciation variant (full *vs*. reduced); Error bars: 95% confidence intervals; Control participants' accuracy in bold horizontal lines (left).



The participants were significantly less accurate on the NTTs than on the TTs. They were also significantly less accurate on the reduced than the full tokens. This difference was significant for the TTs (β=0.38, SE=0.16, z=2.32, $p<0.05$) and the NTTs (β=1.52, SE=0.18, z=8.22, $p<0.0001$). Importantly, there was also a significant interaction between training and reduction: the participants performed significantly worse on reduced NTTs.

Table 2: Statistical model for the response accuracy to target items (first occurrence). Standard error is indicated by SE.

Fixed effects		β	SE	t	p<
(intercept)		0.93	0.30	3.05	0.0001
No Training		-0.99	0.36	-2.73	0.001
Reduced		-0.41	0.17	-2.49	0.05
No Training : Reduced		-1.05	0.24	-4.34	0.0001
Random effects			**variance**	**SD**	
Item	Intercept		1.21	1.10	
Participant	Intercept		1.12	1.06	

3.2. Reaction times to Target words (TTs and NTTs)

The participants responded faster to the reduced than the full TTs while the reverse was true for the NTTs (*cf.* Figure 1, right panel). We tested the significance of these RT results (*cf.* table 3), using as fixed control predictors Stimulus duration and RT to the previous trial, which indicates the participant's local speed in the test. We found that participants reacted significantly faster if they had also answered fast to the previous trial, and that they reacted more slowly to long stimuli, due to the fact that RTs were measured from word onset. More importantly for our research questions, reduction of the stimuli was not significant as a fixed predictor, while it was significant as a random slope on Item, which means that the effect of reduction varied depending on the word type. In addition, we found a significant interaction of training and reduction, showing that training only affected the reduced forms (β= 0.06, SE=0.03, z=1.97, $p<0.05$) but not the full forms (β= 0.01, SE=0.03, z=0.51, n.s.).

Table 3: Statistical model for the response times to the correctly responded target items (first occurrence). Standard error is indicated by SE.

Fixed effects		β	SE	t	p<
(intercept)		5.43	0.49	11.06	0.0001
RT preceding trial		0.08	0.02	3.41	0.0001
Stimulus duration		0.18	0.07	2.61	0.01
No Training		-0.02	0.03	-0.82	n.s.
Reduced		-0.05	0.03	-1.67	n.s.
No Training : Reduc		0.09	0.04	2.44	0.05
Random effects		**variance**	**SD**	**Corr**	
Item	Intercept	0.005	0.07		
	Reduced	0.007	0.09	-0.47	
Participant	Intercept	0.010	0.10		
Residual		0.031	0.18		

4. DISCUSSION

This study examined the effect of implicit training of Dutch low-proficiency learners on French schwa alternation. The participants' accuracy to the taught schwa words (TTs) in the lexical decision task was relatively good (67%) and significantly better than that of the non-Trained schwa words (NTTs; 43%), showing that the participants had learnt the TTs.

Our first research question was whether learners can comprehend reduced and full forms equally well when they have been exposed to these forms equally often. The participants' accuracy scores suggest that this is the case because we found a difference in accuracy score between the two forms of only 6 percent points. This difference in accuracy is, however, significant. The small advantage of the full over the reduced form could be explained by the phonetic differences between the two: reduced and full forms differ in duration, number of phonemes and consonant clusters (many reduced forms begin with illegal consonant clusters in French). Possibly, the learners had more trouble encoding the more complex and often illegal onsets. Also, shorter forms leave less time for processing while containing the same amount of information.

The second question this study addressed is: do the participants generalize the schwa alternation pattern to previously encountered words? The large and significant difference between our participants' accuracy scores for the full and reduced NTTs, which were not statistically different from that of the twelve control participants, show that the participants did not generalize the schwa alternation pattern to these previously known words. The RT data is in line with this conclusion as it shows different patterns for the TTs and the NTTs. The participants responded faster to the reduced than to the full TTs, suggesting similar processing times for the two forms given that RTs were measured from word onset and that reduced forms were shorter. Conversely, participants had longer RTs to the reduced than to the full NTTs, suggesting more processing difficulty for the reduced NTTs. Participants were thus not able to extract a general pattern about alternating schwa words and probably stored both the full and reduced forms of the particular words on which they were trained. Possibly, participants failed to generalize because they needed more examples or also examples for words they already know in order to deduce a generalization. It could also be that learners need a night of sleep [15] to integrate the new pattern. Finally, the learnt words may not have been phonologically similar enough to the known words. Note, however, that the TTs also showed large phonological variation, which should have been a sufficient cue for the participants that the reduction pattern applied to all kinds of words.

All these results have direct implications for L2 teaching. It is easy and relatively cheap to provide learners with more exposure to reduced word forms through audio or video recordings, for example. Moreover, if generalization of reduction patterns is not easily achieved by beginner learners, teachers should try to use or present both forms for the maximum of words in the classroom.

In conclusion, our study has shown that beginning learners' difficulties with reduced word forms can be largely overcome when learners are equally exposed to both the reduced and full forms of the words. However, implicit training on one day is not enough to trigger the generalization of schwa alternation processes to previously encountered words.

5. ACKNOWLEDGEMENTS

This work was supported by a VICI grant (277-70-010) from the Netherlands Organisation for Scientific Research (NWO) to the second author.

6. REFERENCES

[1] Ernestus, M., Warner, N. 2011. An introduction to reduced pronunciation variants [Editorial]. *Journal of Phonetics* 39(SI), 253-260.

[2] Nouveau, D. 2012. Limites perceptives du e caduc pour les apprenants néerlandophones. *The Canadian Journal of Applied Linguistics* 15(1), 60-78.

[3] Chan, R., Leung, J. 2014. Implicit Learning of L2 Word Stress Regularities. *Second Language Research* 30 (4), 463–84.

[4] Hansen, A. 1994. Etude du E caduc – stabilisation en cours et variations lexicales. *Journal of French Language Studies* 4, 25-54.

[5] Hanique, I., Ernestus M., Schuppler B. 2013. Informal Speech Processes Can Be Categorical in Nature, Even If They Affect Many Different Words. *The Journal of the Acoustical Society of America* 133 (3), 1644-55.

[6] Racine, I. 2008. Les Effets de l'effacement du Schwa sur la production et la perception de la parole en français. PhD Thesis. University of Geneva. http://archive-ouverte.unige.ch/unige:602.

[7] *Franconville* 2004. Frans voor havo/vwo, Livre de textes 3, Tweede druk, Utrecht/Zutphen, , Thieme Meulenhoff,.
Grandes Lignes 2008. 1 Livre de textes A, Groningen/Houten, Nordhoff.

[8] http://carrefour-education.qc.ca/multimedia/recherche, http://commons.wikimedia.org/wiki/Accueil, http://maxint.mpi.nl/repos/experiment_stimuli/visual/picdbase/picdbase.html

[9] Boersma, P., Weenink, D. 2001. Praat, a System for Doing Phonetics by Computer. *Glot International* 5, 341-45.

[10] Peirce, J. 2007. PsychoPy-Psychophysics Software in Python. *Journal of Neuroscience Methods* 162 (122), 8–13.

[11] Presentation version 13.0. Neurobehavioral Systems, www.nbs.com

[12] Jaeger, F. 2008. Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of Memory and Language* 59 (4), 434–46.

[13] Baayen, R., Davidson, D., Bates, D. 2008. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* 59 (4), 390–412.

[14] R Development Core Team. 2007. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org

[15] Tamminen, J., Payne J., Stickgold R., Wamsley E., Gaskell G. 2010. Sleep Spindle Activity Is Associated with the Integration of New Memories and Existing Knowledge. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (43), 14356–60.